08/817673

Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

94307843.6

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

A.G. POELS

Den Haag,den
The Hague, 20/11/95
La Haye,le

Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.:
Demande n°:
94307843.6

Anmeldetag:
Date of filing:
Date de dépôt:
25/10/94

Anmelder:
Applicant(s):
Demandeur(s):

BRITISH TELECOMMUNICATIONS public limited company
London EC1A 7AJ
UNITED KINGDOM

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:

Voice-operated services

In Anspruch genommene Prioriät(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:

G10L5/06

Bemerkungen:
Remarks:
Remarques:

VOICE-OPERATED SERVICES

The present invention is concerned with automated voice-interactive services employing speech recognition, particularly, though not exclusively, for use over a telephone network.

A typical application is an enquiry service where a user is asked a number of questions in order to elicit replies which, after recognition by a speech recogniser, permit access to one or more desired entries in an information bank. An example of this is a directory enquiry system in which a user, requiring the telephone number of a telephone subscriber, is asked to give the Town and Road of the subscriber's address, and the subscriber's surname.

A speech recognition apparatus comprising a store of data containing entries to be identified and information defining for each entry a connection with a word of a first set of words and a connection with a word of a second set of words;
speech recognition means; and control means operable:

(a) so to control the speech recognition means as to identify by reference to stored recognition information for the first set of words one or more words of the first set which meet a predetermined criterion of similarity to first received voice signals;

(b) upon such identification, to compile a list of all words of the second set which are defined as connected with entries defined as connected also with the identified word(s) of the first set; and

(c) so to control the speech recognition means as to identify by reference to stored recognition information for the second set of words one or more words of the list which resemble(s) second received voice signals. Other aspects of the invention are defined in the claims.

Some embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings.

The embodiment of the invention now to be described addresses the same directory enquiry task as was discussed in the introduction. In this case however one proceeds at the first recognition operation by retaining as "possible candidates" two or more possible towns. The subsequent recognition of the reply to the road question then proceeds by reference to stored data to pertaining to all road names which exist in any of the candidate towns. Similarly, the name recognition stage employs recognition data for all candidate roads in candidate towns. The number of candidates retained at each stage can be fixed, or (preferably) all candidates having a recognition 'score' above a defined threshold may be retained.

Before describing the process in more detail, the architecture of a directory enquiry system will be described with reference to Figure 1. A speech synthesiser 1 is provided for providing announcements to a user via a telephone line interface 2, by reference to stored, fixed messages in a message data store 3, or from variable information supplied to it by a main control unit 4. Incoming speech signals from the telephone line interface 2 are conducted to a speech recogniser 5 which is able to recognise spoken words by reference to, respectively, town, road or name recognition data in recognition data stores of 6, 7, 8.

A main directory database 9 contains, for each telephone subscriber in the area covered by the directory enquiry service, an entry containing the name, address and telephone number of that subscriber, in text form. The town recognition data store 6 contains, in text form, the names of all the towns included in the directory database 9, along with stored data to enable the speech recogniser 5 to recognise those town names in the speech signal received from the telephone line interface 2. In principle, any type of speech recogniser may be used, but for the purposes of the present description which is assumed that the recogniser 5 operates by recognising distinct phonemes in the input speech, which are decoded by reference to stored data in the store 6 representing a decoding tree structure constructed in advance from phonetic translations of the town names stored in the store 6, decoded by means of a Viterbi algorithm. The stores 7, 8 for road recognition data and name recognition data are organised in the same manner. Although, for example, the name recognition data store 8 contains all the names included in the directory database 9, it is configurable by the control unit 4 to limit the recognition progress to only a subset of the names, typically by flagging the relevant parts of the recognition data so that the "recognition tree" is restricted to recognising only those names within a desired subset of the names.

This enables the 'recognition tree' to be built before the call commences and then manipulated during the call. By restricting the active subset of the tree, computational resources can be concentrated on those words which are most likely to be spoken. This reduces the chances that an error will occur in the recognition process.

The system operation is illustrated by means of the flowchart set out in Figure 2. The process starts (10) upon receipt of an incoming telephone call signalled to the control unit 4 by the telephone line interface 2; the control unit responds by instructing the speech synthesiser 1 to play (11) a message stored in the message store 3 requesting the caller to give the name of the required town. The caller's response is received (12) by the recogniser. The recogniser 3 then performs its recognition process (13) with reference to the data stored in the store 6 and communicates to the control unit 4 the name of the town which most clearly resembles the received reply or (more preferably) the names of all those towns which meet a prescribed threshold of similarity with the received reply. We suppose (for the sake of this example) that four towns meet this criterion. The control unit 4 responds by instructing the speech synthesiser to play (14) a further message from the message data store 3 and meanwhile access (15) the directory database 9 to compile a list of all road names which are to be found in any of those four towns. It then uses (16) this information to update the road recognition data store 7 so that the recogniser 3 is able to recognise only the road names in that list.

The next stage is that a further response is received (17) from the caller and is processed by the recogniser 3 utilising the data store 7; suppose that five road names meet the recognition criterion. The control unit 4 then instructs the playing (19) of a further message asking for the name of the desired telephone subscriber and meanwhile (20) retrieves from the database 9 a list of the names of all subscribers residing in roads having any of the five names in any of the four towns, and updating the name recognition data store 8 in a similar manner as described above for the road recognition data store. Once the user's response is received (22) by the recogniser, the name may be recognised (23) by reference to the data in the name recognition data store.

It may of course be that more than one name meets the recognition criterion; in any event, the database 8 may contain more than one entry for the same name in the same road in the same town. Therefore at step 24 the number of directory entries which have one of the recognised names and one of the recognised roads and one of the recognised towns is tested. If the number is manageable, for example if it is three or fewer, the control means instructs (25) the speech synthesiser to play an announcement from the message data store 3, followed by recitation of the name, address and telephone number of each entry, generated by the speech synthesiser 1 using text-to-speech synthesis, and the process is completed (26). If, on the other hand, the number of entries is excessive then further steps 27 to be discussed further below, will be necessary in order to meet the caller's enquiry.

It will be seen that the process described will have a lower failure rate than a system which chooses only a single candidate town, road or name at each stage of the recognition process, since by retaining second and further choice candidates the possibility of error due to mis-recognition is reduced though there is increased risk of recognition error due to the larger vocabulary. A penalty for this increased reliability is of course increased computation time, but by ensuring that the road and name recognition processes are conducted over only a limited number of the total number of roads and names in the database, the computation can be kept to manageable proportions.

Moreover, compared with a system in which a second-stage recognition is unconstrained by the results of previous recognition (e.g. one where the 'road' recognition processes is not limited to roads in towns already recognised) the proposed system would, when using recognisers (such as those using Hidden Markov Models) which internally "prune" intermediate results, be less liable to prune out the desired candidate favour of other candidate roads from unwanted towns.

It will be seen too, that the number of possible lists will, in most applications, be so large as to prohibit their preparation in advance, and hence the construction of the list is performed as required. Where the recogniser is of the type (e.g. recognisers using Hidden Markov models) which require setting up for a particular vocabulary, there are two options for updating the relevant store to limit the recogniser's operation to words in the list. One is to start with a fully set-up recogniser, and disable all the words not in the list; the other is to clear the relevant recognition data store and set it up afresh (either completely, or by adding words to a permanent basic set).

Generally the first option would be preferred, with the second option being invoked in the case of a short list, or where the data change frequently.

It is assumed in the above description that the recognisers always produce a result - i.e. that the town (etc) or towns which give the nearest match(es) to the received response. It would of course be possible to permit output of a "fail" message in the event that a reasonably accurate match was not found. In this case, as in the case where the number of entries found at step 24 is excessive, further action will be required. This could simply be switching of the call to a manual operator. Alternatively further information may be processed automatically as in the following example which assumes that, following an excessive number of entries recognised, namely to play to the caller a further message asking for an additional reply which may be checked against the existing recognition results. For example the caller may be asked to spell the name of the person required. This may be processed as follows:

- compare the spelled name with the list of names recognised, and if a match is obtained output the matching entry to the user as at step 25;
- in the event of no match, check whether the spelled name matches any of the names compiled at step 20, and if so, output that name(s).

Another possibility at this point would be to repeat the name search 23, but over all the names included in the database, with the same processing as for the spelled name recognition.

The criterion for limiting the number of recognition 'hits' at steps 13, 18 or 23 may be that all candidates are retained which meet some similarity criterion, though other criteria such as retaining always a fixed number of candidates may be chosen if preferred. It may be, in the earlier recognition stages, that the computational load and effect on recognition performances of retaining a large town (say) with a low score is not considered to be justified, whereas retaining a smaller town with the same score might be. In this case the scores of a recognised word may be weighted by factors dependent on the number of entries referencing that word, in order to achieve such differential selection. It is not necessary that the response to be recognised be discrete responses to discrete questions. They could be words extracted by a recogniser from a continuous sentence, for systems which work in this way.

In the examples discussed above, a list of words (such as road names) to be recognised is generated based on the results of an earlier recognition of a word (the town). However it is not necessary that the unit in the earlier recognition step or in the list be single words; they could equally well be sequences of words. One possibility is a sequence of the names of the letters of the alphabet, for example a list of words for a town recognition step may be prepared from an earlier recognition of the answer to the question "please spell the first four letters of the town name." If recording facilities are provided (as discussed further below) it is not essential that the order of recognition be the same as the order of receipt of the replies (it being more natural to ask for the spoken word first, followed by the spelled version, though it is preferred to process them in the opposite sequence).

Another situation in which it may be desired to vary the scope of the speech recogniser's search is where it can be modified on the basis not of previous recogniser results but of some external information relevant to the enquiry. In a directory enquiry system this may be a signal indicating the origin of a telephone call, such as the calling line identity (CLI) or a signal identifying the originating exchange. In a simple implementation this may be used to restrict "town" recognition to those towns located in the same or an adjacent exchange area to that of the caller. In a more sophisticated system this identification of the calling line or exchange may be used to access stored information compiled to indicate the enquiry patterns of the subscriber in question or of subscribers in that area (as the case may be).

For example, a sample of directory enquiries in a particular area might show that 40% of such calls were for numbers in the same exchange area and 20% for immediately adjacent areas. Separate statistical patterns might be compiled for business residential lines, or for different times of day, or other observed trends.

The effect of this approach is to improve the system reliability for common enquiries at the expense of uncommon ones. Such a system thus aims to automate the most common or straightforward enquiries, with other calls being routed to a human operator.

As an example, Figure 1 additionally shows a CLI detector 20, (used here only to indicate the originating exchange) which is used to select from a store 21 a list of likely towns for enquiries from that exchange, to be used by the control unit 4 to truncate the "town" search, as indicated in the flowchart of Figure 3, where the calling line indicator signal is detected at step 10a, and selects (12a) a list of towns from the store 21 which is then used (12b) to update the town recognition store 6 prior to the town recognition step 13. The remainder of the process is not shown as it is the same as that given in Figure 2.

In the above described embodiment, no account is taken of the relative probability of recognition, for example if the town recognition step 13 recognises town names Norwich and Harwich, then when, at road recognition step 18 the recogniser has to evaluate the possibility that the caller said "Wright Street" (which we suppose to be in Norwich) or "Rye Street" (in Harwich), no account is taken of the fact that the spoken town bore a closer resemblance to "Norwich" than it did to "Harwich". If desired however, the recogniser may be arranged to produce (in known manner) figures or "scores" indicating the relative similarity of each of the candidates identified by the recogniser to the original utterance and hence the supposed probability of it being the correct one. These scores may then be fed as a priori probabilities to the next recognition stage to bias the selection. This may be implemented in the process depicted in Figure 2 as follows.

Step 13. The recogniser produces for each town, a score - e. g.

Harwich 40%
Norwich 25%
Nantwich 20%
Northwich 15%

Step 15. When the road list is compiled the appropriate score is appended to the road name, e. g.

Wright Street 25%
Rye Street 50%
North Street (assumed to exist in both Norwich and Nantwich) 40%
:

and stored in the store 7.

Step 18. When the recogniser comes to recognise the road name, it weights its own scores by the scores from the store 7 before producing a final result. For example if the recogniser would have assigned scores of 60% and 30% to High Street and Rye Street then the weighted scores would be

Wright Street (Norwich) 25% x 60% = 15
Rye Street (Harwich) 50% x 30% = 15
North Street 40% x 10% = 4

Similar modification would of course occur for the steps 20, 21, 23. This is just one example of a scheme for score propagation.

The possibility of switching to a manual operation in the event of a "failure" condition has already been mentioned. However, further automated steps may be taken under such conditions.

A failure condition can be identified by noting low recogniser output "scores", or of excessive numbers of recognised words all having similar scores (whether by reference to local scores or to weighted scores). Such a condition may arise in an unconstrained search like that of the town recognition of step 13 in Figure 24 in which case it may be that better results might be obtained by performing (for example) the road recognition step first and compiling a list of all towns containing the roads found, to constrain a subsequent town recognition step. Or it may arise in a constrained search such as that of step 13 in Figure 3 or steps 18 and 23 in Figure 24, where perhaps the constraint has removed the correct candidate from the recognition set; in the case removing the constraint - or applying a different one - may improve matters.

Thus one possible approach is to make provision for recording the caller's responses, and in the event of failure, reprocessing them using the steps set out in Figure 2 (except the "play message" steps 11, 14, 19) but with the original sequence town/road/name modified. There are of course six permutations of these. One could choose that one (or more) of these which experience shows to be the most likely to produce an improvement. The result of such a reprocessing could be used alone, or could be combined with the previous result, choosing for output those entries identified by both processes.

Another possibility is to perform an additional search omitting one stage, and comparing the results as for the 'spelled input' case.

If desired, processing using two (or more) such sequences could be performed routinely (rather than only under failure conditions); to reduce delays an additional sequence might commence before completion of the first; for example (in Figure 4) an additional, unconstrained "road" search 30 could be performed (without recording the road name) during the "which name" announcement. From this, a list of names is compiled (31) and the name store updated (32). Once the names from the list have been recognised (33) a town list may be compiled (34) and the town store updated (35). Then at step 36 the spoken town name, previously stored at step 37 may be recognised. The results of the two recognition processes may then be compiled, suitably be selecting (38) those entries which are identified by both processes. The remaining steps shown in Figure 4 are identical to those in Figure 2.

CLAIMS

1. A speech recognition apparatus comprising a store of data containing entries to be identified and information defining for each entry a connection with a word of a first set of words and a connection with a word of a second set of words;
speech recognition means; and control means operable:

(a) so to control the speech recognition means as to identify by reference to stored recognition information for the first set of words one or more words of the first set which meet a predetermined criterion of similarity to first received voice signals;

(b) upon such identification, to compile a list of all words of the second set which are defined as connected with entries defined as connected also with the identified word(s) of the first set; and

(c) so to control the speech recognition means as to identify by reference to stored recognition information for the second set of words one or more words of the list which resemble(s) second received voice signals.

2. A speech recognition apparatus according to Claim 1 in which the speech recognition means is operable to identify a plurality of words of the first set.

3. A speech recognition apparatus according to Claim 2, in which the speech recognition means is operable upon receipt of the first voice signal to generate for each identified word a measure of similarity with the first voice signal, and the control means is operable to generate for each word of the list a measure obtained from the measure(s) for the relevant word(s) of the first set, and the speech recognition means is operable upon receipt of the second voice signal to perform the identification of one or more words of the list in accordance with a

recognition process weighted in dependence on the measures generated for the words of the list.

3. A speech recognition apparatus according to claim 1, 2 or 3 in which, the apparatus includes a store containing recognition data for all words of the second set and the control means is operable following the compilation of the list and before recognition of the word(s) of the list to mark in the recognition data store those items of data therein which correspond to the words not in the list or those which correspond to words which are in the list, whereby the recognition means may ignore all words so marked or, respectively, not marked.

4. A speech recognition apparatus according to claim 1, 2 or 3 in which, the apparatus includes a store containing recognition data for all words of the second set and the control means is operable following the compilation of the list and before recognition of the word(s) of the list to generate recognition data for each word of the list and place it in the recognition store.

5. A speech recognition apparatus according to any one of the preceding claims in which the control means is operable to select for output that entry or entries defined as connected both with an identified word(s) of the first set and an identified word of the second set.

6. A speech recognition apparatus according to any one of claims 1 to 4 in which the store of data also contains information defining for each entry a connection with a word of a third set of words, and the control means is operable:

(d) to compile a list of all words of the third set which are defined as connected with entries each of which is also defined as connected both with an identified word of the first set and an identified word of the second set; and

(e) so to control the speech recognition means as to identify by reference to stored recognition information for the third set of words one or more words of the list which resemble(s) third received voice signals.

7. A speech recognition apparatus according to any one of the preceding claims including means to store at least one of the received voice signals, the apparatus being arranged to perform an additional recognition process in which the control means is operable:

(a) so to control the speech recognition means as to identify by reference to stored recognition information for the second set of words a plurality of words of the second set which meet a predetermined criterion of similarity to the first received voice signals;

(b) to compile an additional list of all words of the first set which are defined as connected with entries defined as connected also with the identified words of the second set; and

(c) so to control the speech recognition means as to identify by reference to stored recognition information for the first set of words one or more words of the said additional list which resemble(s) the second received voice signals.

8. A speech recognition apparatus according to Claim 7 including means to recognise a failure condition and to initiate the said additional recognition process only in the event of such failure being recognised.

9. A telephone information apparatus according to any one of the preceding claims, further comprising a telephone line connection; and means responsive to receipt via the telephone line connection of signals indicating the origin

of a telephone call to access stored information identifying a subset of the said set of words and to restrict to that subset the operation of the speech recognition means for that set.

10. A telephone information apparatus comprising a telephone line connection; a speech recogniser for recognising spoken words received via the telephone line connection, by reference to recognition data representing a set of possible utterances; and means responsive to receipt via the telephone line connection of signals indicating the origin of a telephone call to access stored information identifying a subset of the set of utterances and to restrict the recogniser operation to that subset.

11. A speech recognition apparatus comprising:

a store defining a first set of words;

a store defining a second set of words;

a store containing entries to be identified;

a store containing information relating each entry to a word of the first set and to a word of the second set;

speech recognition means operable upon receipt of a first voice signal to identify one or more words of the first set;

means to generate a list of all words of the second set which are related to an entry to which an identified word of the first set is also related; and

speech recognition means operable upon receipt of a second voice signal to identify one or more words of the list.

12. A recognition apparatus comprising:

a store defining a first set of patterns;

a store defining a second set of patterns;

a store containing entries to be identified;

a store containing information relating each entry to a pattern of the first set and to a pattern of the second set;

recognition means operable upon receipt of a first input pattern signal to identify one or more patterns of the first set;

means to generate a list of all patterns of the second set which are related to an entry to which an identified pattern of the first set is also related; and recognition means operable upon receipt of a second input pattern signal to identify one or more patterns of the list.

13. A speech recognition apparatus comprising:

(i) a store of data containing entries to be identified and information defining for each entry a connection with a signal of a first set of signals and a connection with a word of a second set of words;

(ii) means for identifying a received signal as corresponding to one of the first set;

(iii) control means operable to compile a list of all words of the second set which are defined as connected with entries defined as connected also with the identified signal of the first set; and

(iv) speech recognition means operable to identify by reference to stored recognition information for the second set of words one or more words of the list which resemble(s) received voice signals.

14. A speech recognition apparatus according to Claim 13 in which the first set of signals are voice signals representing spelled versions of the words of the second set or initial portions thereof and the identifying means are formed by the speech recognition means operating by reference to stored recognition information for the said spelled voice signals.

15. A method of identifying entries in a store of data by reference to stored information defining connections between entries and words, comprising:

(a) identifying one or more of the said words as present in received voice signals;

(b) compiling a list of those of the said words defined as connected with entries defined as connected also with the identified word(s);

(c) identifying one or more of the words of the list as present in the received voice signals.

16. A speech recognition apparatus comprising:

a) a store of data containing entries to be identified and information defining for each entry a connection with at least two words;

b) a speech recognition means able to identify by reference to stored recognition information for a defined set of words at least one word or word sequence which meets some predefined criterion of similarity to a received voice signal;

(c) a control means operable:

i) to compile a list of words which are defined as connected with entries defined as connected with a word previously identified by the speech recognition means; and

ii) so to control the speech recognition means as to identify by reference to stored recognition information for the compiled list one or more words or word sequences which resemble a further received voice signal.

ABSTRACT

A speech recognition system accesses a database such as a telephone directory where entries are linked to words of two or more vocabularies (town, street, name). A first voice signal produces a number of candidate results (e.g. towns) which are used to identify entries and compile a list of words in the second vocabulary (e.g. streets) to which those entries are also linked. The list then forms a vocabulary for recognition of a second voice signal.

Fig 2.

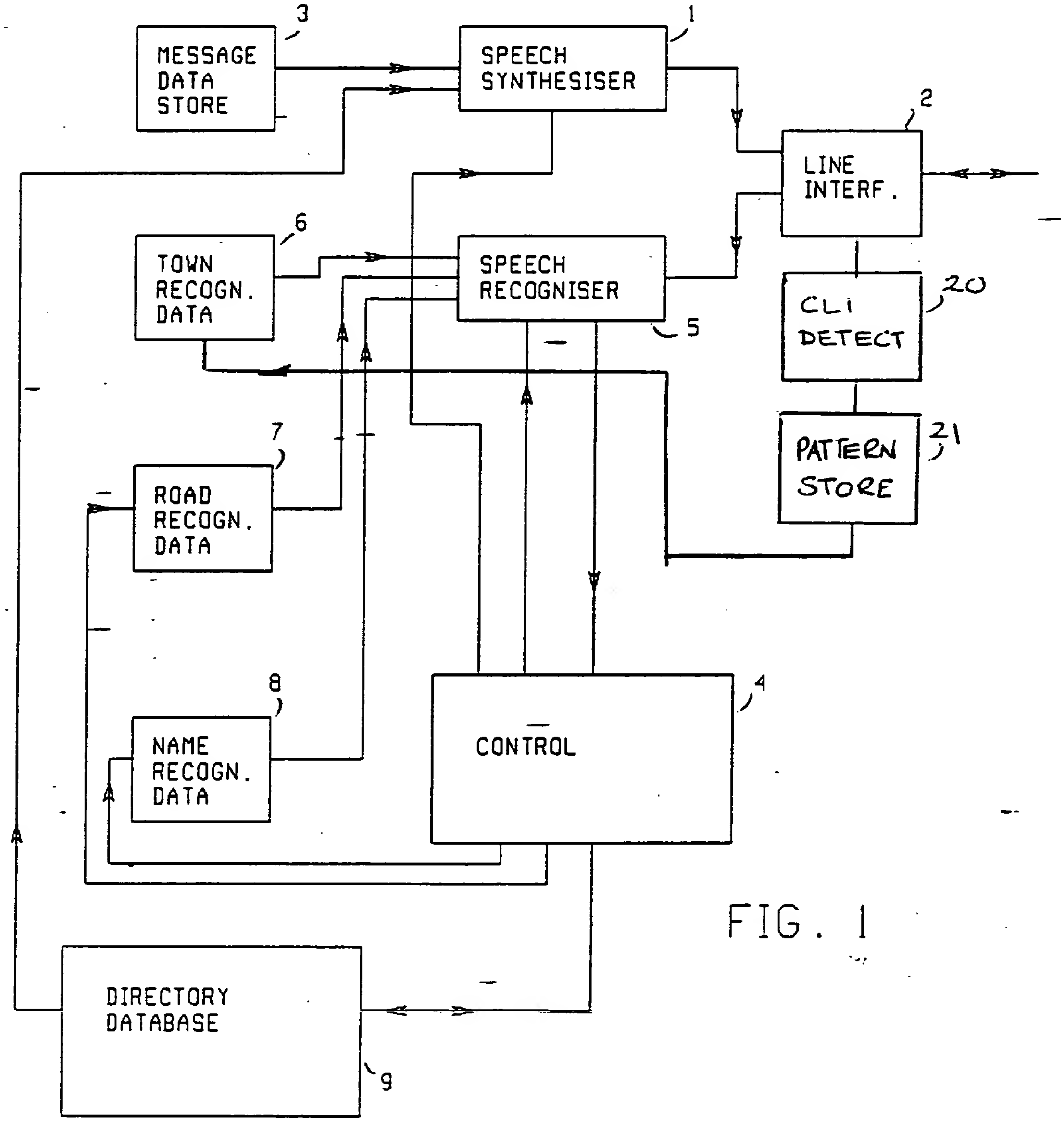
MESSAGE DATA STORE
3
SPEECH SYNTHESISER
1
LINE INTERF.
2
TOWN RECOGN. DATA
6
SPEECH RECOGNISER
5
CLI DETECT
20
PATTERN STORE
21
ROAD RECOGN. DATA
7
NAME RECOGN. DATA
8
CONTROL
4
DIRECTORY DATABASE
9

FIG. 1

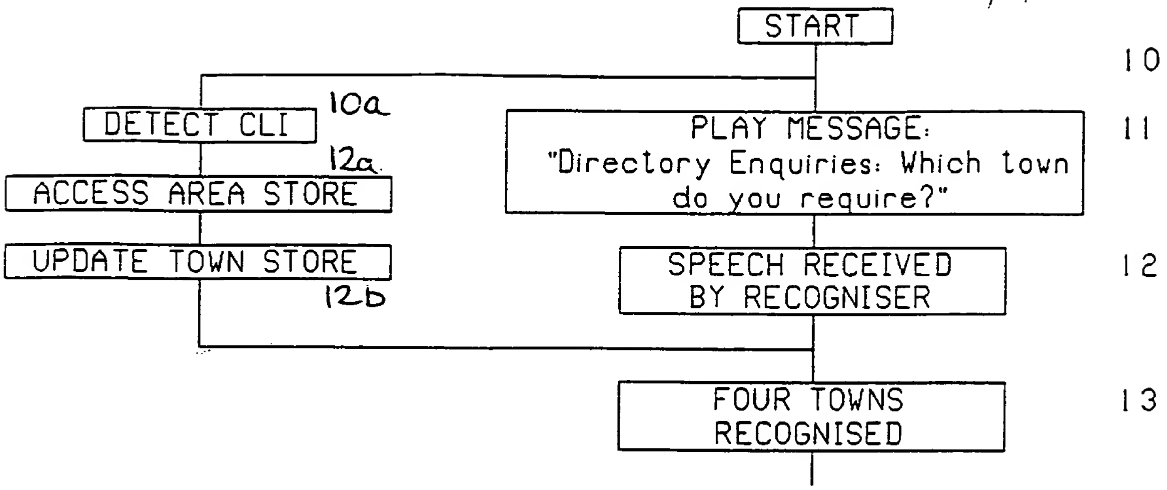
START
10
DETECT CLI
10a
PLAY MESSAGE: "Directory Enquiries: Which town do you require?"
11
12a
ACCESS AREA STORE
UPDATE TOWN STORE
12b
SPEECH RECEIVED BY RECOGNISER
12
FOUR TOWNS RECOGNISED
13

FIG. 3

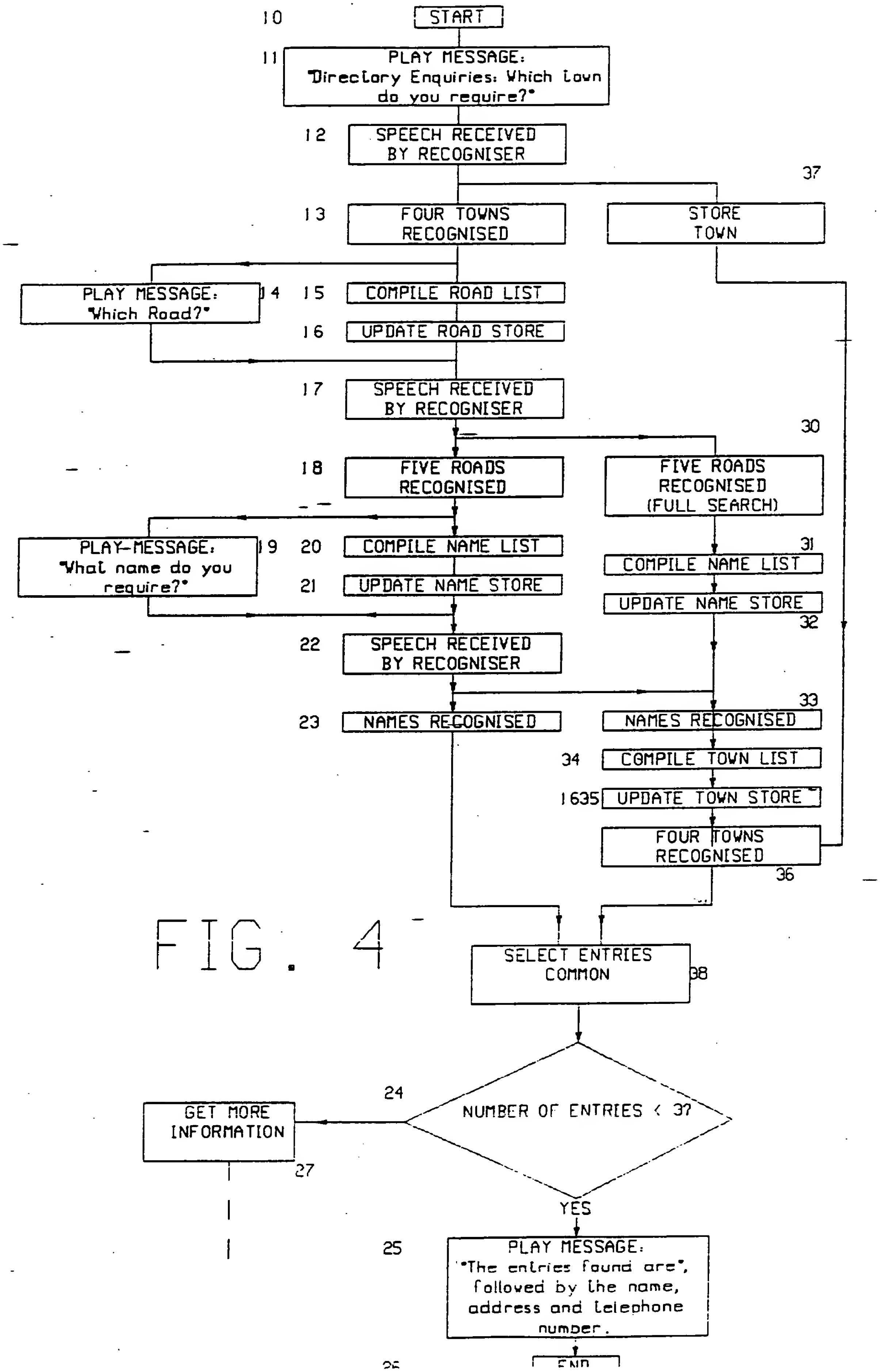
10
START
11
PLAY MESSAGE:
"Directory Enquiries: Which town do you require?"
12
SPEECH RECEIVED BY RECOGNISER
37
STORE TOWN
13
FOUR TOWNS RECOGNISED
PLAY MESSAGE:
"Which Road?"
14
15
COMPILE ROAD LIST
16
UPDATE ROAD STORE
17
SPEECH RECEIVED BY RECOGNISER
30
18
FIVE ROADS RECOGNISED
FIVE ROADS RECOGNISED (FULL SEARCH)
PLAY MESSAGE:
"What name do you require?"
19
20
COMPILE NAME LIST
31
COMPILE NAME LIST
21
UPDATE NAME STORE
UPDATE NAME STORE
32
22
SPEECH RECEIVED BY RECOGNISER
33
23
NAMES RECOGNISED
NAMES RECOGNISED
34
COMPILE TOWN LIST
1635
UPDATE TOWN STORE
FOUR TOWNS RECOGNISED
36
SELECT ENTRIES COMMON
38
24
GET MORE INFORMATION
NUMBER OF ENTRIES < 3?
27
YES
25
PLAY MESSAGE:
"The entries found are", followed by the name, address and telephone number.
26
END

FIG. 4

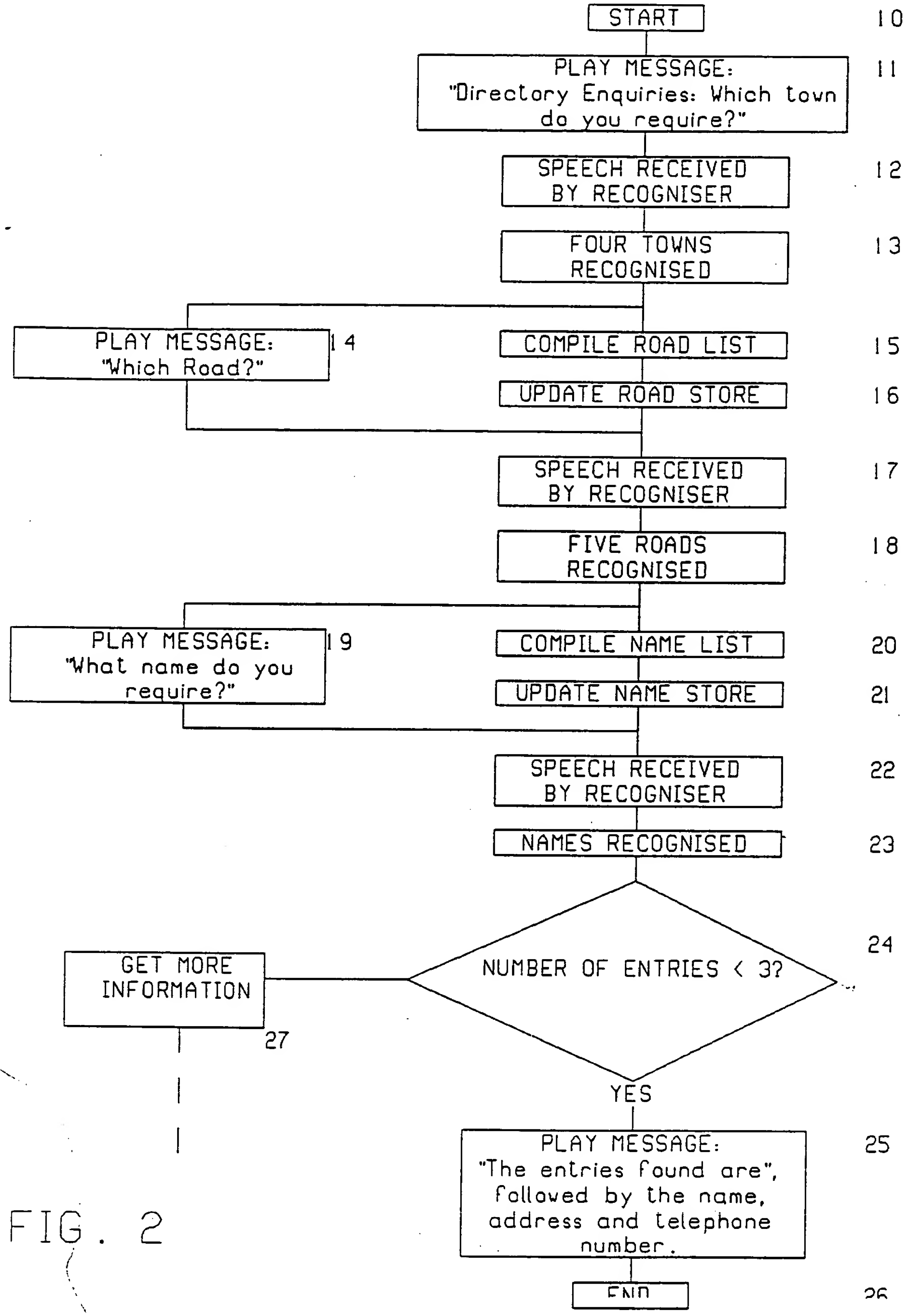
START
10
PLAY MESSAGE: "Directory Enquiries: Which town do you require?"
11
SPEECH RECEIVED BY RECOGNISER
12
FOUR TOWNS RECOGNISED
13
PLAY MESSAGE: "Which Road?"
14
COMPILE ROAD LIST
15
UPDATE ROAD STORE
16
SPEECH RECEIVED BY RECOGNISER
17
FIVE ROADS RECOGNISED
18
PLAY MESSAGE: "What name do you require?"
19
COMPILE NAME LIST
20
UPDATE NAME STORE
21
SPEECH RECEIVED BY RECOGNISER
22
NAMES RECOGNISED
23
NUMBER OF ENTRIES < 3?
24
GET MORE INFORMATION
27
YES
PLAY MESSAGE: "The entries found are", followed by the name, address and telephone number.
25
END
26

FIG. 2